(19) Federal Republic of Germany
German Office of Patents and Trademarks

(10) **DE 103 58 333 A1** 2005.07.14

(12) # Document of Disclosure

(21) Reference: **103 58.333.5**
(22) Date of registration: **Dec.12, 2003**
(43) Date of disclosure: **July 14, 2005**

(51) Int Cl.[7]: **H04M 3/22**
G10L 17/00

(71) Registering company
Siemens AG, 80333 Munich, Germany

(72) Inventors:
Schmöller, Franz, 82110, Germering, Germany
Zimmermann, Beate, 82065, Baierbrunn, Germany

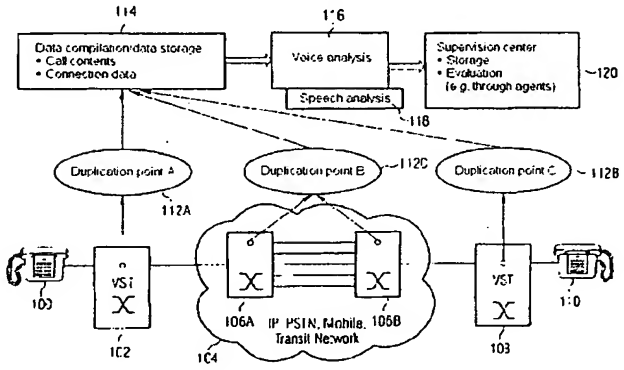
(56) Documents taken into account for judging the possibility of patenting:
US2002/01 50 096 A1
US2001/00 43 687 A1
US2001/00 36 821 A1
US 59 13 161 A
US 64 96 483 B1
US 64 24 701 B1

The following details have been taken from the documents submitted by the registering company

The application for examination is submitted according to § 44 PatG.

(54) Designation: **Procedure, Network Element, and Network Arrangement for the Supervision of Telecommunications**

(57) Summary: The invention applies to the field of legal supervision and compilation of relevant data by the prosecution authorities, generally summarized through the term of Lawful Interception (LI). Well-known LI solutions are always based on endpoint addresses, which lead practically to incomplete LI data and/or to the supervision of a large number of users not subjected to the LI action. According to the invention, a process is planned for the supervision of telecommunications, where a supervision region is selected and supervision parameters are given in the form of speech or voice characteristics. In the supervision region, the entire telecommunications traffic is recorded and submitted to speech and voice analysis, and the portion of the telecommunications traffic matching the supervision parameters is selected for composing supervision data.



Description

[0001] Modern communications networks offer many speech communications possibilities for users through various media. As in the past, much use is made of speech services of classical line telephone networks, often called a fixed network or PSTN. Mobile telephone extensions are also widely used, and with the increase of efficient Internet connections, packet-based speech connections gain an increasing importance.

[0002] The large number of media and protocols of present and future communications networks makes it easier for users with criminal intentions to plan and arrange criminal actions, while at the same time the legal supervision and compilation of relevant data by the prosecuting authorities, generally summarized through the term of Lawful Interception (LI), is made considerably more difficult.

State of Technology

[0003] Until now, LI solutions have always been based on endpoints, a state which requires that for a successful LI action, all the endpoints of a user to be supervised must be known. For instance, the following details must be known: All Internet providers, all fixed network extensions, all mobile endpoints etc. that the user employs. Often, this requirement cannot be fulfilled. For instance, the user can evade LI action through public pay phones or mobile prepaid cards, which are often issued without any identity control or with insufficient identity control, or by employing Internet access that is open to the public, e.g. in Internet cafés or libraries.

[0004] A second problem exists, when an LI action based on endpoints takes up in principle also users, for whom the required LI action is not applicable and for whom a legal supervision order is not applicable, e.g. for family members. This is especially difficult when the endpoints to be supervised include a public extension, because in this case, the LI action inevitably affects uninvolved users.

Nature of the Task

[0005] The invention's purpose is therefore to name a process, a network element, and a network arrangement that avoid the disadvantages mentioned above.

Implementation Example

[0006] This task is implemented by means of a process for the supervision of telecommunications, where the supervision region is selected in the form of speech or voice characteristics. Within the supervision region, the entire telecommunications traffic is recorded and subjected to speech or voice analysis, and the portion of the telecommunications traffic matching the supervision parameters is selected for composing supervision data.

[0007] Supervision regions can be selected (automatically) as geographical or logical regions. In an advantageous form, the selection of a geographical supervision region is done by first recording a roaming profile for a user and then selecting the supervision region according to the limits of the user's roaming field. To this end, a mobile roaming region of the user can be selected. In an additional form, a logical supervision region is defined through at least one address space, e.g. through the address space of a specific Internet provider.

[0008] Speech or voice characteristics of a user or of a group of users can be selected as supervision parameters, where supervision parameters are selected in such a way that each user can be definitely identified.

[0009] Alternatively, instead of supervising individual users, the invention can be employed in a geographical region that contains a higher potential for disturbance, in order to analyze all telecommunications. Speech or voice characteristics are chosen as supervision parameters, which result in recognition of higher aggression potential among all the users operating in the supervision region. In this case user identification is not necessary; the speech or voice evaluation can be based, for instance, on statistical supervision parameters, e.g. sound level, pitch of voice, speech velocity etc.

[0010] In order to fulfill legal requirements and/or diminish the machine computation requirements for speech or voice analysis, telecommunications traffic can be placed in intermediate storage at first and be later evaluated by speech and voice analysis. In this case, instead of placing the telecommunications traffic in intermediate storage, the special case of buffering can also be provided, if the legal requirements demand this. Technical means can be provided in this case to prevent access to buffered, unfiltered, telecommunications traffic, so that no (possibly inadmissible) general supervision takes place in the supervision region.

[0011] The invention foresees also a network element for the supervision of telecommunications, which provides the following:

- Means for receiving telecommunications traffic indicated by a duplication point, where the telecommunications traffic includes connection contents and data referring to the traffic;
- Means for analyzing the connection contents on the basis of supervision parameters containing at least speech or voice characteristics;
- Means for forwarding the portion of the telecommunications traffic matching the supervision parameters to a supervision center.

[0012] In addition, the network element can have storage means for the described buffering of the telecommunications traffic.

[0013] A preferred network element for the supervision of data following a Voice over Internet Protocol contains the following.

- Packet filters for recognizing a Voice over Internet Protocol process;
- Means for intermediate storage of the data stream in the Voice over Internet Protocol process;
- Means for decoding packet-based speech data and for creating a sound stream;

- Means for comparing the sound stream with supervision parameters showing at least speech and voice characteristics;
- Means for conducting the intermediately stored data stream to the supervision center responding to the matching of the sound stream with the supervision parameters.

[0014] Additionally, in the network element prepared according to the invention, there can be means for speech recognition in order to convert the relevant telecommunications into text, as well as means to forward the text to the supervision center.

[0015] The network element prepared according to the invention can be integrated in switching elements in order to reduce the quantities of data by filtering them at the place of their creation.

[0016] The invention concerns also a network arrangement for telecommunications supervision, which includes the following:

- Access switching elements, which supply telecommunications services to terminal devices;
- A connection network;
- One or more duplicating points to duplicate the telecommunications traffic and to forward the duplicated traffic to a supervision network element constructed according to the invention.

[0017] The advantage of the invention is that at first in the supervision region the full telecommunications are taken into account for an LI action, are put to intermediate storage or are buffered, they are then analyzed according to speech or voice characteristics, which when applied makes it possible to implement the destination-oriented and complete supervision of a user or of a certain user group in the supervision region and prevents at the same time the supervision of noninvolved users.

[0018] In another case of application, the invention serves to ascertain an aggression potential. In this action, the evaluation of the aggression potential can take place also in parallel to the first case of application (supervision of selected users).

[0019] The invention can also be employed together with the conventional LI process, where the supervision is based on addresses of the user to be supervised, in order to forward only the telecommunications traffic produced really by the user to be supervised up to a work station in the supervision center and not to the telecommunications traffic created by users to whom the LI action does not apply. With this combination, the inclusion of public endpoints is not complicated. At the same time, the expense for the required speaker recognition is low, because the supervision region is accordingly small (it contains then the endpoints selected for the conventional LI process).

[0020] In the following, an example of implementing the invention is described in detail together with three drawings.

[0021] Fig. 1 shows a network arrangement for the supervision of telecommunications in a schematic diagram.

[0022] Fig. 2 shows schematically the course of the supervision of telecommunications in real time.

[0023] Fig. 3A-B shows schematically the course of the supervision of telecommunications with intermediate storage or buffering of the telecommunications.

[0024] Fig. 1 shows a typical network arrangement with two users, **100** and **110**, given as an example, who are provided through local exchanges, **102** and **108**, with telecommunications services. The two exchanges are connected through a transit network **104**, which has transit exchanges **106A-B**.

[0025] The transit network **104** can be a conventional PSTN transit network, a mobile telephone transit network, a transit network based on the Internet Protocol IP, a transit network based on the Ethernet, or any other transit network. Instead of using the classical local exchanges **102** and **108**, the users **100** and **110** can also be supplied with telecommunications services through Voice over IP servers, mobile radio base stations, and other means of network access, which have not been illustrated.

[0026] Fig. 1 shows also three duplication points **112A-C**, where telecommunications data are duplicated. In this process, both telecommunications contents (e.g. call contents and data) and control information (e.g. protocol-specific signaling information and connection data) are duplicated.

[0027] The duplicated data is forwarded to a first component **114** for data compilation. The first component **114** receives the duplicated data from the duplicating points **112** and controls these points. If there are several duplicating points **112** as in the given example, which duplicate the data referring to the same connection (here referring to the connection between the users **100** and **110**), the first component **114** selects a suitable duplicating point **112** in order to record the data of this connection. Otherwise, the first component **114** receives continuously data from all duplication points **112** and rejects data recorded twice.

[0028] As already explained, the telecommunications traffic can be put at first to intermediate storage by the first component **114**, in order to comply with the legal requirements and/or in order to reduce the computation requirements for means of speech or voice analysis.

[0029] From the first component **114**, the recorded telecommunications data is forwarded to a second component **116** for voice analysis. The voice analysis is carried out with supervision parameters, which represent the voice and speech characteristics of a user to be supervised. These supervision parameters can be obtained from existing speech records containing the speech of the user to be supervised. If there are no speech records of this kind, suitable speech records can be produced by supervising a communications endpoint that can be assigned to the user, who is to be supervised, by means of a classical LI action. In order to prevent the falsification of the supervision parameters by other users of the same endpoint, an LI agent can examine the speech records before they are transformed into supervision parameters.

[0030] The second component **116** can be complemented by a third component **118** for speech recognition (speech-to-text transformation). This is advantageous in combination with the second component **116**, because the analysis of the speech data, which is necessary for the speaker identification provides intermediate results, which can be further used for the speech identification. The speech identification transforms the speech data assigned to the user, who is to be supervised, into machine-readable texts.

[0031] Only the telecommunications data that can be assigned to the user, who is to be supervised, is forwarded to a supervision center **120** (called also Monitoring Center, MC). There, the data is stored and evaluated, e.g. by an LI agent or automatically.

[0032] In this action, the supervision is limited to one supervision region. Various considerations can lead to the specification of a supervision region:

- The specification of the supervision region can take place purely administratively; e.g. the entire (geographical) territory of a country can be specified as a supervision region.
- The supervision region can be restricted to a certain access provider or to his (logical) communications address spaces, if it is known that the user utilizes only this provider. The same can be applied to two or more providers utilized by the user.
- The supervision region can be found automatically by supervising for a certain time an endpoint ready for roaming of the user to be supervised and by determining the supervision region on the basis of the (geographical) roaming region (e.g. as the established roaming region extended by a safety radius). Mobile roaming regions and Internet accesses ready for roaming are especially suitable. In the case of the Internet, the geographical data of IP addresses can be recorded alternatively, from where certain services that can be assigned to the user were called up (e.g. an Email query or Login for Online Banking) in order to create a roaming profile.
- A suitable algorithm can combine the above-mentioned criteria in order to supervise for instance only part of a roaming region situated in the territory of a certain country, because perhaps outside this country, LI measures are subject to other requirements. Likewise, by an "OR" combination of the geographical region with the logical region, a larger supervision region can be defined.

[0033] In Fig. 1, three duplication points **112A-C** are displayed. In practice, only one of these duplication points is necessary. In this process, a duplication point **112A-B** can be arranged in conjunction with a local exchange **102, 108**, or it can be realized centrally in conjunction with the transit network **104** (duplication point **112C**).

[0034] While it is possible to duplicate the entire traffic of the communications network with one duplication point **112C** coupled to the transit network **104**, the requirements of such a duplication point **112C** are quite high. Therefore, it might be advantageous to install duplication points **112A-B** at several peripheral points in the communications network. Depending on the hierarchy and structure of the communications network, the direct subscriber access elements (e.g. local exchanges) or network elements of a higher hierarchical level are offered.

[0035] In both cases, before the speech and/or voice analysis, a pre-selection in the second component **116** is possible on the basis of subscriber addresses (e.g. telephone numbers, IP addresses) carried out for instance by the first component **114** or by the duplication points **112**. The pre-selection can serve to keep up the specified limits of the geographical or logical supervision region. Moreover, the pre-selection through positive lists can basically exclude certain addresses from the supervision and/or the pre-selection through negative lists can basically subject certain addresses to the supervision.

[0036] In this way, an advantageous combination of known LI processes referring to endpoints utilized by the user to be supervised can be achieved with the invention, when for instance the telecommunications traffic produced at the endpoints to be supervised will then be subjected to filtering with automatic speech identification. In other words, the specification of the supervision region can take place through telecommunications addresses with the criteria applicable to current LI actions.

[0037] Various aggregations and integration levels of the elements and components presented in Fig. 1 are possible technically and advantageous; however, these aggregations and integration levels are not legally admitted in all countries. In order to reduce the large data quantities during the supervision of the entire telecommunications traffic already at their origin to the relevant data, it would be desirable to implement a combination of duplication point **112C**, first component **114** and second component **116** directly in connection with one of the transit nodes 106A-B of the transit network **104**. If this is not possible due to legal causes, at least a separate implementation of the above-mentioned combination should be made as near as possible to the data origin, e.g. transit nodes, with a respectively efficient interface.

[0038] A schema of a possible form of the process according to the invention is shown in Fig. 2. In the example presented, the supervision takes place with respect to the connection. At first, a connection set-up **200** is examined to see, whether it is relevant for the LI action (step **202**). In step **202**, the LI relevance is examined with regard to criteria established earlier, e.g. an examination is carried out regarding the question, whether the connection origin or destination belongs to the supervision region.

[0039] If the connection set-up LI is relevant, a query **204** can be provided showing, whether the LI action is carried out on the basis of DN subscriber numbers (DN = Directory Number). If this occurs [point (a) of the process], the agreement of the subscriber numbers (of the calling and called subscriber) with the LI criteria is examined (step **206**); this is branched in step **208** to point (1) of the process in case of non-agreement.

[0040] If the query in step **202** is negative, i.e. if the connection set-up is not LI-relevant, a decision is made in step **222** according to specifications, whether for the present supervision region, the aggression potential of the connected subscribers should be examined. If the query in step **222** is negative [point (1) of the process], step **224** is examined, showing whether the data comes from a delayed data analysis (post-processing, see below with reference to **Fig. 3**) If yes, the post-processing is finished. If no, the LI connection is finished and the subscriber connection can be continued without LI influence.

[0041] If however, query **208** or **222** is fulfilled, or if query **204** is not fulfilled [point (b) of the process], the speech or voice analysis (step **210**) is reached. If the detected parameters, which characterize the current speaker, agree with the supervision parameters at least for one speaker (step **212**), then in step **214** [point (2) of the process] the forwarding of telecommunications data to the supervision center MC (step **216**) is initiated; all the data and contents related to this connection are registered after a positive identification of a user to be supervised without any further examination. In addition, the telecommunications traffic transferred until the secured positive identification can be put to intermediate storage - this is not displayed. In case of a positive identification, this section of a call preceding the identification is forwarded also to the supervision center; otherwise, it is rejected.

[0042] If in step 214, no agreement could be established between one of the speakers and one of the users to be supervised, a disposition analysis takes places with regard to the aggression potential. If aggressions are recognized, e.g. by means of keywords in the evaluation of a speech recognition (element 118 from Fig. 1) and/or according to the tone position, speech velocity, and similar factors, additional measures can be initiated. In the given example, branching is carried out to point (2) of the process in order to forward the data to the supervision center MC (step 216). In the supervision center, with the amount of connections established in a certain region with a higher aggression potential, a rising disturbance for instance can be found, and further measures can be initiated, e.g. an intensified visual observation in the region or a strengthening of the security forces.

[0043] However, if no aggressions can be derived from the disposition analysis in steps 218 and 220, branching can be carried out to point (1) of the process (see above). An alternative way is to return to point (b) of the process in order to examine, whether a speaker to be supervised has joined the connection, e.g. when the connection is enlarged to a conference or when a user actually to be supervised asks at first unsuspicious users to establish the connection in order to join this connection later on.

[0044] With regard to Fig. 3A-B, the course will be described below for the process according to the invention, if intermediate storage of the telecommunications must be provided. As mentioned already, both technical and legal backgrounds can be decisive for an intermediate storage of the telecommunications traffic followed by the evaluation in post-processing. In these cases, the real time process given in Fig. 2 is not directly applicable.

[0045] Fig. 3A shows the course for the storage of the telecommunications traffic. A connection set up in step 300 is duplicated in all data, both call data CC (CC = Call Content) and supervision relevant data IRI (IRI = Interception Related Information; e.g. date of call, duration, etc.), through duplication points 112 (step 302) and is put to intermediate storage (step 304), before the connection is terminated regularly in step 306.

[0046] Fig. 3B shows the post-processing which can take place subsequently. Due to the real time requirements, which do not exist in this case, the post-processing can for instance be carried out also during times of lower traffic. For the post-processing, data are taken in step 310 from the intermediate storage, and if an analysis of this data is foreseen (query 312), it will be checked in step 314, whether the LI action takes place based on the subscriber addresses DN. If yes, the processing will be continued in point (a) of the process from Fig. 2. If no, it will be checked in step 316, whether a speech and/or voice analysis should take place. If yes, the processing will be continued point (b) of the process from Fig. 2; otherwise, the post-processing is terminated.

[0047] A special case of voice analysis for the purpose of speaker recognition results, if any protocol of Voice over IP (VoIP) is employed, e.g. H.323, SIP, or proprietary derivates such as Net2Phone with the respective signaling and/or control protocols. In this matter, a distinction is made between two case: The voice analysis with the purpose of speaker identification takes place directly in the duplication point, which is also called a 'trial, or in a downstream analysis component, as shown for instance in Fig. 1.

[0048] For speaker identification in the duplication point, a filter is used to recognize VoIP traffic, e.g. a Berkeley Packet Filter (BPF), which works for instance with IP addresses or TCP/UDP port numbers. The VoIP traffic is decomposed into signaling information and useful information (payload), e.g. H.225 (signaling channel), H.245 (control channel), and RTP (RTP = Real Time Protocol, which forwards the payload). The signaling information or control information is often called meta-data.

[0049] The supervision parameters characterizing the user/users to be supervised are stored in the duplication point or can be retrieved through the duplication point. In the duplication point, the control and/or signaling information is decoded and the payload stream is extracted. The Codec used for speech coding is determined. If necessary, the payload stream and the associated control and/or signaling information is buffered, and a temporary sorting of the packets of the payload stream takes place.

[0050] As it is not possible, in general, to produce directly the speech parameters from the VoIP-coded speech data for comparison with the supervision parameters, using the determined Codec the VoIP-coded speech data is initially transformed into spoken speech, into a sound stream, or into an intermediate format. The sound stream is then analyzed like spoken speech (see above), and when a user to be supervised is positively identified, the telecommunications traffic to be assigned to this VoIP stream including the data buffered until then is forwarded to the supervision center.

[0051] However, if the analysis of the VoIP traffic is downstream, possibly as displayed in Fig. 1, the above-mentioned functions are implemented in a suitable component 116. In this case, the duplication point serves only to filter/identify and forward the VoIP traffic to this component for voice/speaker analysis.

[0052] Naturally, it is also possible in connection with VoIP traffic and even necessary under the circumstances already mentioned to put at first the VoIP traffic to intermediate storage and then analyze it for speaker identification.

[0053] The filtering in the duplication point can also include for VoIP traffic a filtering according to preset source and/or destination addresses in order to restrict the LI action to a geographical or logical region or to certain addresses employed generally by the user.

[0054] As shown already, the invention can be used on the one hand in the framework of known LI actions that are implemented on the basis of known communications addresses of users to be supervised, and can be completed for the purpose of identifying among the possible users of a communications endpoint those users who really are to be supervised and to forward only their communications to a supervision center, and on the other hand in order to analyze the entire communications traffic in a region that can be easily defined and to forward only calls of users to be supervised to a supervision center.

[0055] In the given examples of implementation, the invention was described as an example only with regard to fixed network communications and VoIP communications. Basically, the invention is suitable for supporting the supervision of telecommunications of all kinds. Therefore, the invention is not restricted to the implementation example but includes all forms of telecommunications, e.g. speech connections between two subscribers, conferences with any number of subscribers, connections to announcement systems or

answering equipment over any telecommunications systems and protocols.

Patent Claims

1. Process for the supervision of telecommunications including the following steps:

- Selection of a supervision region;
- Specification of supervision parameters in the form of speech or voice characteristics;
- Compilation of the entire telecommunications traffic in the supervision region;
- Implementation of a speech and voice analysis of the recorded telecommunications traffic; and
- Selection of the portion of the telecommunications traffic matching the supervision parameters for composing supervision data.

2. Process according to Claim 1, by which as supervision region, a geographical supervision region is selected, where at first a roaming profile is recorded for a user and the supervision region is selected on the basis of the limits of the roaming region of the user.

3. Process according to Claim 2, by which a geographical supervision region is selected on the basis of the limits of a mobile roaming region of the user.

4. Process according to Claim 1, which as supervision region defines a logical supervision region defined by means of at least one address space.

5. Process according to one of Claims 1 to 4, which as supervision parameters selects the speech or voice characteristics of a user or of a group of users, where supervision parameters are selected in such a way that each user can be definitely identified.

6. Process according to one of Claims 1 to 3, which as supervision region selects a geographical region selected, that contains a higher disturbance potential and speech or voice characteristics are selected as supervision parameters allows the determination of a higher aggression potential among all users operating in the supervision region.

7. Process according to one of Claims 1 to 6, by which after the step for compiling the telecommunications traffic, a step for intermediate storage of the telecommunications traffic follows and by which the speech and voice analysis is implemented downstream for the telecommunications traffic put to intermediate storage.

8. Network element (**116**) for telecommunications supervision including the following:

- Means for receiving telecommunications traffic indicated by a duplication point (**112**), where the telecommunications traffic includes connection contents and data referring to the connection;
- Means for analyzing the connection contents on the basis of supervision parameters containing at least speech and voice characteristics;
- Means for forwarding the portion of the telecommunications traffic matching the supervision parameters to a supervision center (**120**).

9. Network element (**116**) according to Claim 8, which includes also means of storage (**114**) for the intermediate storage of the telecommunications traffic.

10. Network element (**116**) according to Claim 8 for the supervision of data according to a process of Voice over Internet Protocol, where the network (**116**) includes the following:

- Packet filters for recognizing a process of the Voice over Internet Protocol;
- Means for intermediate storage of the data stream in the process of the Voice over Internet Protocol;
- Means for decoding packet based speech data and for creating a sound stream;
- Means for comparing the sound stream with supervision parameters showing at least speech and voice characteristics;
- Means for forwarding the intermediately stored data stream to the supervision center responding to matching of the sound stream with the supervision parameters

11. Network element (**116**) according to one of Claims 8 to 10, which includes also means for speech recognition (**118**) as well as means for forwarding the portion converted into text of the telecommunications traffic matching the supervision parameters to the supervision center (**120**).

12. Switching element (**102, 106, 108**) of a telecommunications network, in which a network element (**116**) is integrated according to one of Claims 8 to 11.

13. Network arrangement for telecommunications supervision including the following:

- Access switching elements (**102, 108**), which supply telecommunications services to terminal devices (**100, 110**);
- A connection network (**104**);
- One or several duplicating points (**112**) in order to duplicate the telecommunications traffic and to forward the duplicated traffic to supervision network element (**116**) according to one of Claims 8 to 11.

There follow drawings on 3 pages

Attached drawings

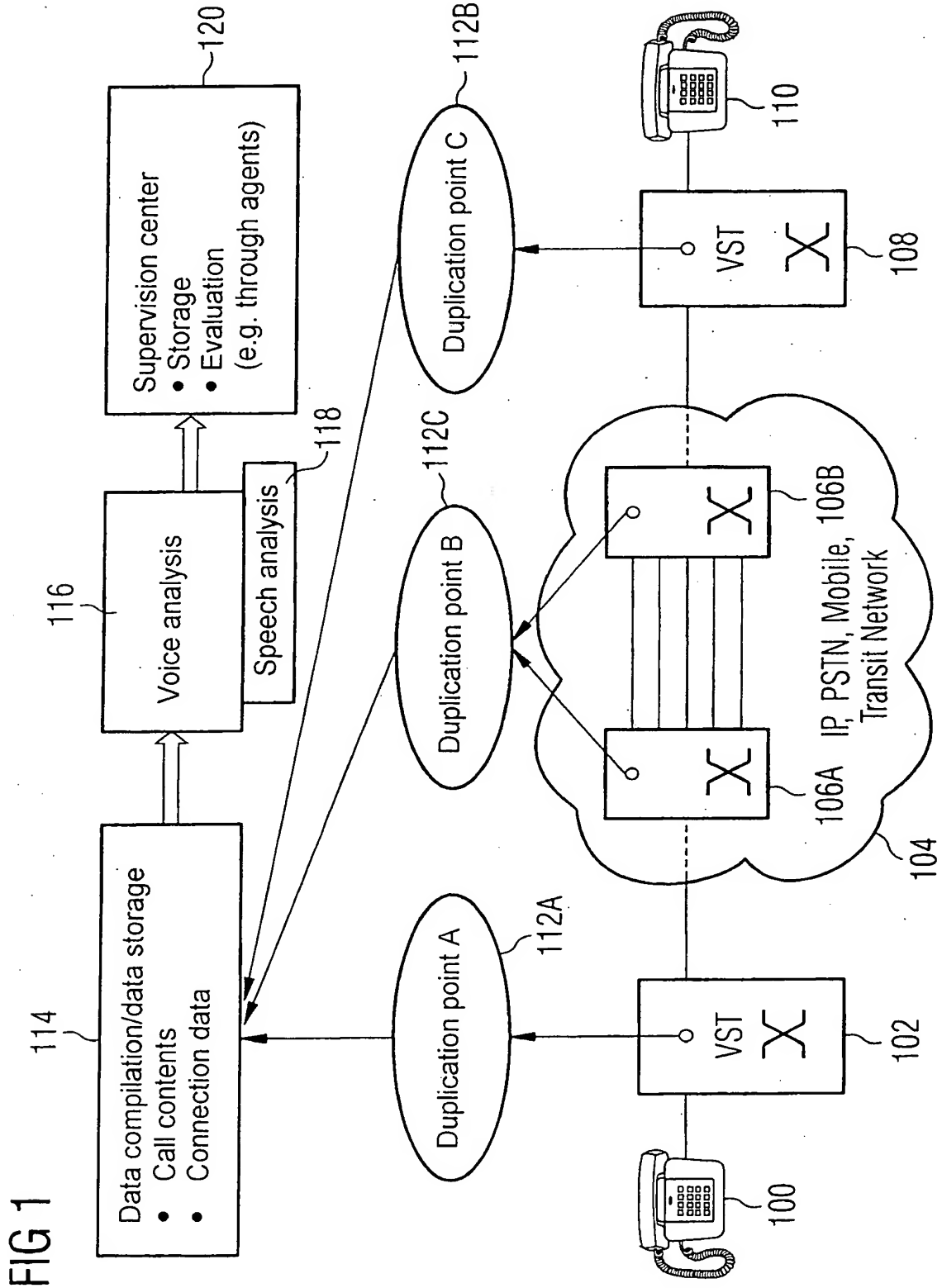


FIG 1

FIG 2

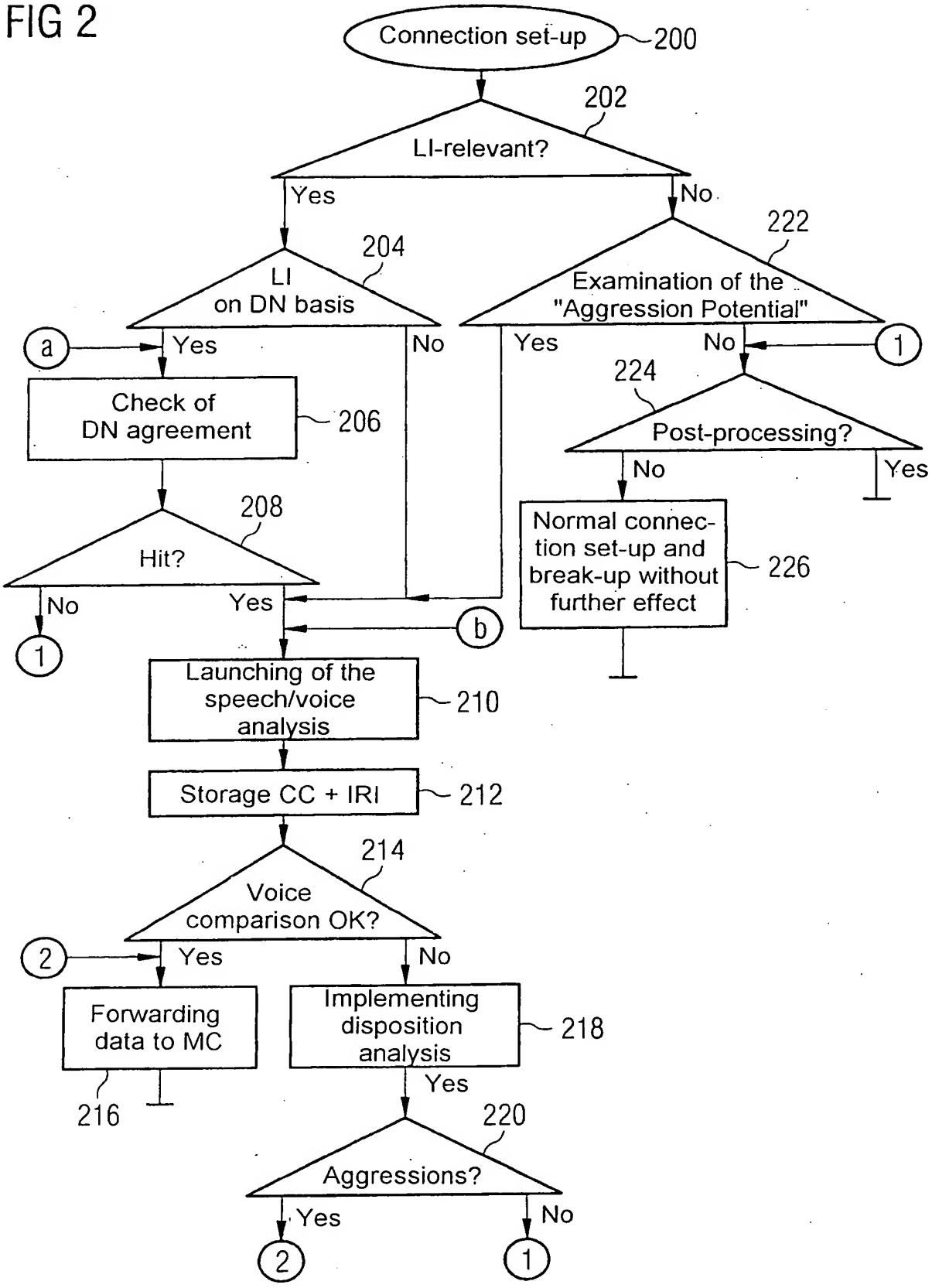

Connection set-up
200
202
LI-relevant?
Yes
No
204
LI
on DN basis
222
Examination of the
"Aggression Potential"
a
Yes
No
Yes
No
1
Check of
DN agreement
206
224
Post-processing?
No
Yes
208
Hit?
Normal connec-
tion set-up and
break-up without
further effect
226
No
Yes
1
b
Launching of the
speech/voice
analysis
210
Storage CC + IRI
212
214
Voice
comparison OK?
2
Yes
No
Forwarding
data to MC
Implementing
disposition
analysis
218
216
Yes
220
Aggressions?
Yes
No
2
1

FIG 3A

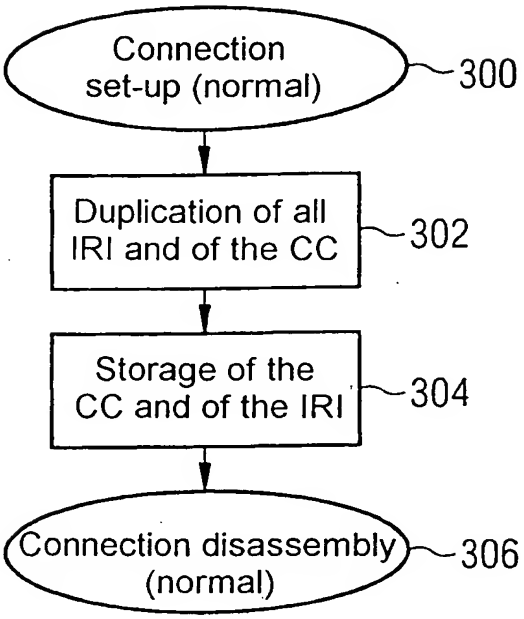
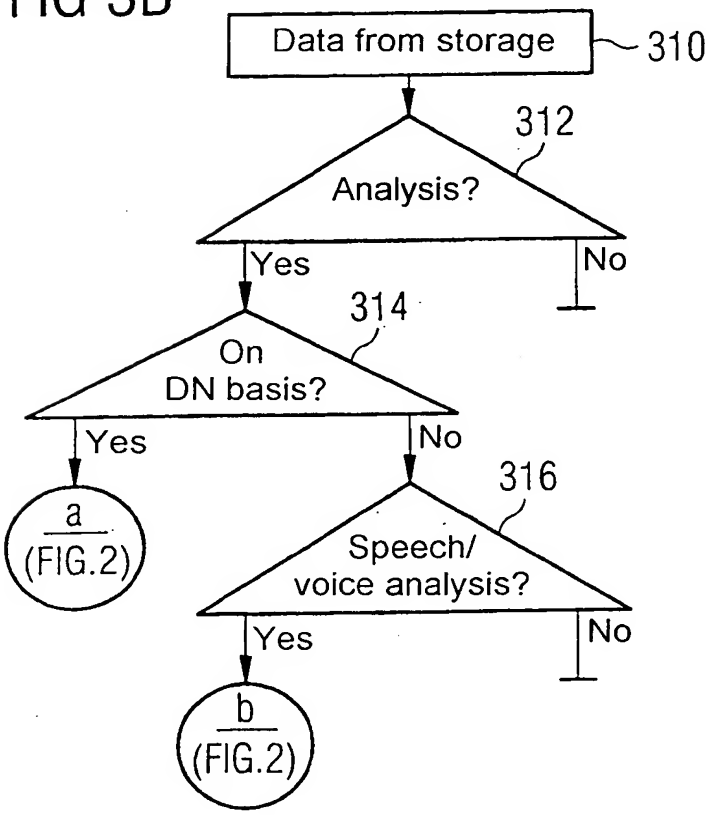

Connection set-up (normal)
300
Duplication of all IRI and of the CC
302
Storage of the CC and of the IRI
304
Connection disassembly (normal)
306

FIG 3B


Data from storage
310
312
Analysis?
Yes
No
314
On DN basis?
Yes
No
316
Speech/ voice analysis?
Yes
No
a
(FIG.2)
b
(FIG.2)